Newcastle University e-prints

Date deposited:

23rd November 2009

Version of file:

Uncorrected proof

Peer Review Status:

Not known

Citation for published item:

U. C. Alsmark;T. Sicheritz-Ponten;P. G. Foster;R. P. Hirt;T. M. Embley. Horizontal gene transfer in eukaryotic parasites: a case study of Entamoeba histolytica and Trichomonas vaginalis. *Methods Mol Biol* 2009,**532** 489-500.

Further information on publisher website:

Publishers copyright statement:

The original publication is available at www.springerlink.com

Use Policy:

Chapter 28

Horizontal Gene Transfer in Eukaryotic Parasites: A Case Study of *Entamoeba histolytica* and *Trichomonas vaginalis*

Cecilia Alsmark, Thomas Sicheritz-Ponten, Peter Foster, Robert Hirt, and Martin Embley

Abstract

Over the past few years it has become apparent that horizontal gene transfer (HGT) has played an important role in the evolution of pathogenic prokaryotes. What is less clear is the exact role that HGT has played in shaping the metabolism of eukaryotic organisms. The main problems are the reliable inference of HGT on a genomic scale as well as the functional assignment of genes in these poorly studied organisms. We have screened the completed genomes of the protists *Entamoeba histolytica* and *Trichomonas vaginalis* for cases of HGT from prokaryotes. Using a fast primary screen followed by a conservative phylogenetic approach, we found 68 and 153 recent cases of HGT in the respective organisms. The majority of transferred genes that fall into functional categories code for enzymes involved in metabolism. We found a broad range of prokaryotic lineages represented among the donors, but organisms that share similar environmental niches with *E. histolytica* and *T. vaginalis*, such as the gut and the vaginal mucosa, dominate.

Key words: *Entamoeba histolytica*, *Trichomonas vaginalis*, genome-wide analysis, phylogeny, metabolic genes, donor lineages, sampling.

1. Introduction

HGT plays a significant role in prokaryotic genome evolution, contributing up to ~20% of the content of a given genome *(1)*. HGT thus provides an efficient means of gaining new phenotypes, such as resistance to antibiotics and new physiological and metabolic capabilities, permitting or facilitating adaptation to new ecological niches *(2–4)*. More recently, data from microbial eukaryotes suggest that HGT also plays a role in

Maria B. Gogarten et al. (eds.), *Horizontal Gene Transfer: Genomes in Flux, vol. 532*

DOI 10.1007/978-1-60327-853-9_28 Springerprotocols.com

eukaryotic genome evolution, particularly among protists that can feed on bacteria *(1, 5–10)*. The sequencing of the genomes of two anaerobic parasites of humans, *Entamoeba histolytica (10)* and *Trichomonas vaginalis (11)*, has recently been completed. The availability of these data provides an excellent opportunity to investigate the role of HGT in the evolution of parasitic protists and in shaping eukaryotic genomes more generally. *Trichomonas vaginalis*, an anaerobic, flagellated parabasalid, is the most common, non-viral, sexually transmitted parasite of humans in the developed and developing world, mainly affecting women by colonising the urethra and the vaginal mucosa *(12)*. *Entamoeba histolytica* is a microaerophilic phagotrophic parasite living in the human gut *(13)*, an environment that is rich in microorganisms and where HGT is thought to be common between bacteria *(14)*. These two human parasites belong to distinct phylogenetic groups and provide ideal model systems for investigating prokaryote-to-eukaryote HGT.

2. Methods

2.1. Primary Screen

We set up a screening pipeline to detect HGT from a prokaryote donor to the ancestors of *T. vaginalis* and *E. histolytica*, respectively. For all analyses we used a previously published suite of python scripts and modules called PyPhy *(15)*, which we modified and adapted for large eukaryotic genome projects (SpyPhy, http://www.cbs.dtu.dk/staff/thomas/pyphy/spyphy.html). As many genes appear to be present in multiple copies, we first carried out a cluster analysis to identify putative paralogues using Blastclust (length coverage threshold $= 0.9$, score average threshold $= 75$, and this length coverage threshold must be met for both neighbours) from the BLAST package *(16)*. The initial proteomes were thus reduced to a number of clusters or singletons at the level of $\geq 75\%$ identity over $\geq 90\%$ of the gene length for cluster members. A single randomly chosen representative of each cluster was used as a seed for BLASTP searches against a database obtained by merging TREMBL and Swissprot. Sequences above an empirically determined given cutoff value (40% similarity over a minimum of 70% of the protein length), and all cluster members were aligned using ClustalW *(17)*. The resulting alignments were automatically processed (allowed gap positions = half, minimum length of a block = 2, maximum number of contiguous non-conserved positions = 20) using GBLOCKS *(18)* to remove sequence positions where the inference of positional homology was ambiguous. Bootstrap (100 replicates) consensus p-distance trees were made using a PAUP pipeline from edited alignments of all proteins, for which there were sufficient

homologues (≥ 3) in SwissProt and TrEMBL to make trees. The trees were analysed to identify those trees where the nearest neighbour to the input gene was a prokaryote. As an additional primary screen for putative HGT, we identified all proteins for which a prokaryote was the top BLAST hit. These initial screens identified about 1000 candidate HGT for each organism. After inspection of alignments, BLAST outputs, tree support values and sequence identities, cases of potential HGT were retained for more detailed phylogenetic analyses. During the manual inspection we discarded short versions of genes already represented, genes that show similarity to a prokaryote orthologue due to a biased amino-acid composition (for example, LLR-repeats) and genes where only a short domain shows similarity to a prokaryote gene.

2.2. Secondary Screen

Each candidate for HGT was analysed by MrBayes *(19)* using the WAG matrix, a gamma correction for site rate variation and a proportion (pinvar) of invariable sites. The analyses were run for 600,000 generations and sampled every 100 generations, with the first 2000 samples discarded as a conservative burn-in. A consensus tree was made from the remaining samples. Because posterior probabilities, the support values used by Bayesian analysis to indicate confidence in groups, have been criticised *(20)*, we also use bootstrapping to provide an additional indication of support for relationships. Each data set was bootstrapped (100 replicates) and used to make distance matrices under the same evolutionary model as in the Bayesian analysis, using custom (P4) software *(21)*. Trees were made from the distance matrices using FastME *(22)* and a bootstrap consensus tree made using P4.

2.3. Evaluation of Trees

Our two-step screen is aimed at detecting recent HGT from a prokaryote donor to an ancestor of *Entamoeba* or *Trichomonas*. By recent HGT, we include cases where the query genes from *T. vaginalis* or *E. histolytica* were clustered inside well-supported prokaryotic groups and/or were separated from any other eukaryotes in the Bayesian tree, by at least two well-supported (posterior probabilities ≥ 0.95, bootstrap ≥ 70) nodes. In cases when tree topologies were more weakly supported but still suggested a possible HGT, we scrutinised bootstrap partition tables for partitions where the query sequence clustered with another eukaryote. If no such partitions were found we considered that gene also to be a putative HGT. We also considered a case to be a putative HGT if no other eukaryote contained the gene in question as determined by BLAST and HSSP scores falling below our thresholds (see below). We focused on recent HGT because they represent the most robust, least controversial, and easily detected examples

of HGT. Our screen was not designed to detect more ancient gene transfers such as those from the mitochondrial endosymbiont *(23)* or those occurring at the base of major eukaryotic clades (*see* **Chapter 7**).

3. Results

3.1. Prevalence of HGT

For *E. histolytica* a total of 5740 trees were made and 548 of these were selected from the primary screen as representing potential HGT. These included all of the genes that had previously been published as HGT for *Entamoeba (8)*. The 548 candidate genes were then processed through the secondary screen to make better trees using a Bayesian approach. From the 548 trees HGT was inferred in 96 cases using the stringent criteria we applied. The remaining 452 trees are not discussed further here, apart from commenting that their topologies rarely reflected the relationships among taxa which are depicted in the universal rRNA tree of life.

In *Trichomonas vaginalis* we found 153 genes where prokaryote-to-eukaryote HGT, from diverse prokaryotes, is supported either by phylogenetic trees or by the presence of a gene in *Trichomonas* and prokaryotes, but in no other eukaryote sampled. We identified 76 genes where the most straightforward interpretation of trees indicated that HGT from a prokaryote to an ancestor of *T. vaginalis* had occurred. We also identified 77 genes where *Trichomonas* and diverse prokaryotes possess a particular gene, but we were unable to find a convincing homologue among other eukaryotes. To infer putative homology, we used a structure-based criterion called the HSSP score *(24, 25)*. The HSSP approach was designed to identify protein homologues in the so-called twilight zone of 20–35% amino acid identity *(24)*. The HSSP curve plots the number of aligned residues and percent residue identity to identify a length dependent threshold above which homology is likely, based upon comparisons of known homologous protein structures (i.e. true positives). In our analysis we identified eukaryotic and prokaryotic proteins as putative homologues when they showed HSSP scores with a distance of ≥ 10 from this threshold value. At this distance from threshold, the frequency of false positives is effectively zero *(24)*.

3.2. Where Do the Genes Come from?

Based upon our trees certain prokaryote lineages are favoured as potential donors for genes now residing on the *Trichomonas vaginalis* and *Entamoeba histolytica* genomes. Thus, in *Trichomonas*, of 42 unrooted trees where we could identify a nearest neighbour to the *Trichomonas* gene, 15 of these trees placed the *Trichomonas* gene next to a member of the Bacteroidetes phylum

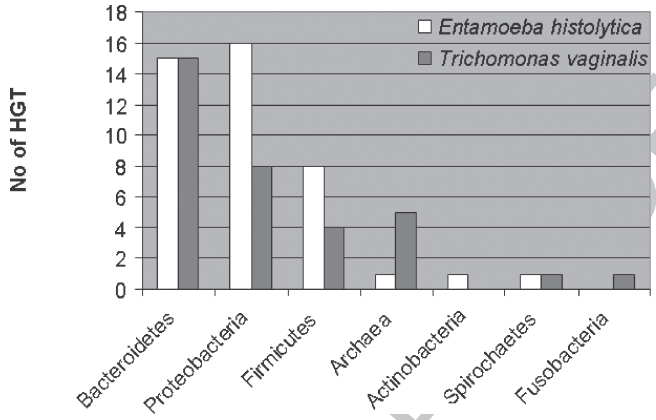

Fig. 28.1. **Taxonomic distribution of putative donors of genes to *Entamoeba histolytica* and *Trichomonas vaginalis***. Among the 76 trees strongly supporting an HGT event from a prokaryotic donor, a total of 42 trees recovered the *Trichomonas vaginalis* sequence(s) as the nearest *neighbour* to a single species or a coherent prokaryotic taxon with $\geq 70\%$ bootstrap support and ≥ 0.95 posterior probability. For *Entamoeba histolytica*, 34 of 68 HGT gave a tree with a nearest neighbour by the same criteria. In both species, the majority of such nearest neighbours comprised one or more members of the Bacteroidetes. Assuming a root outside of the specific *Trichomonas* or *Entamoeba*-prokaryote(s) split, these trees support the notion that *Trichomonas* or *Entamoeba* received the gene from an ancestor shared with that prokaryote lineage. The trees that did not resolve the position of *Trichomonas vaginalis / Entamoeba histolytica* sequences among prokaryotic homologues are not represented in the chart.

(**Fig. 28.1**). A similar bias towards the Bacteroidetes as potential donors was observed **(Fig. 28.1)** for candidate HGT for *Entmoeba histolytica (10)*. In 15 well-resolved trees, *Entamoeba* was recovered next to a member of the Bacteroidetes/Chlorobi group. Bacteroidetes are abundant anaerobic members of the human intestinal flora *(14)* where *Entamoeba* typically resides, and most trichomonads are also associated with the digestive tract and adjacent mucosa *(12)*. For example, Bacteria from the digestive tract are commonly present in the urogenital *(26)* tract where *Trichomonas vaginalis* typically resides. Members of the Bacteroidetes/Chlorobi and *Fusobacterium* (a pathogen colonising the oral cavity) groups are all obligate anaerobes. The observed bias for donor lineages is thus consistent with the idea that anaerobic prokaryotic and eukaryotic cohabitants of the human mucosa are sharing genes *(27)*. **Figure 28.2** shows an intriguing example where *Trichomonas vaginalis* clusters with members of the Bacteroidetes/Chlorobi with maximum support values, and *Entamoeba* clusters with *Fusobacterium*.

3.3. What Kinds of Genes Are Being Transferred?

Most of the transferred genes that can be assigned to a functional category are enzymes involved in metabolism. Furthermore, the genes transferred are diverse and affect many different metabolic pathways; HGT has thus profoundly influenced the

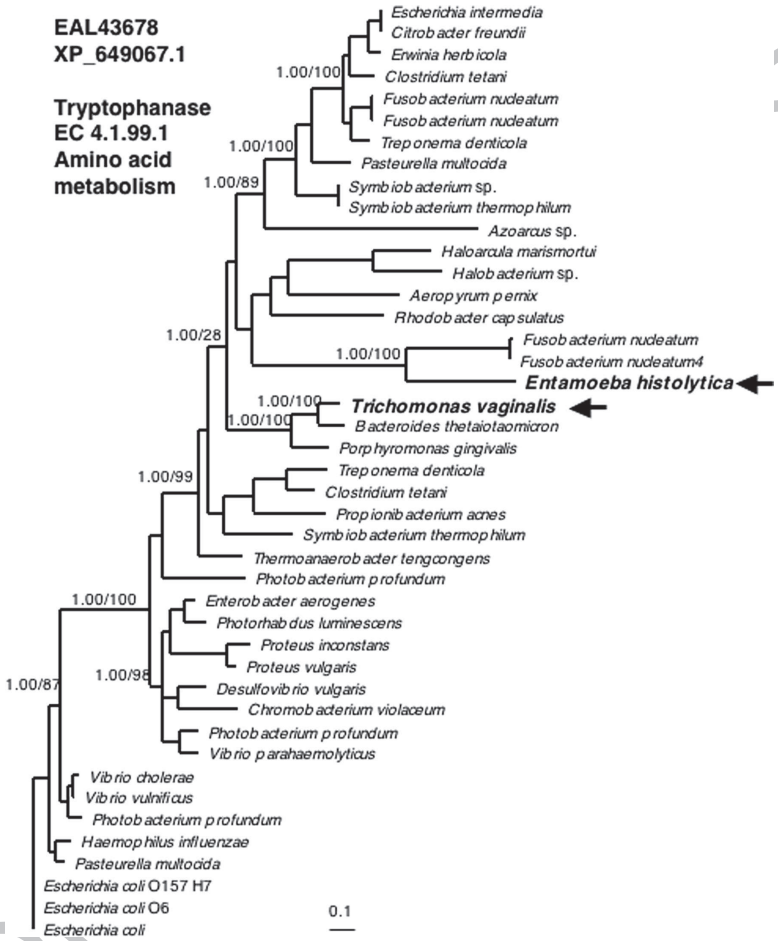

Fig. 28.2. **Example of a phylogenetic tree supporting two strong inferences of HGT**. Consensus MrBayes tree with support values corresponding to the posterior probabilities of the Bayesian analysis followed by the corresponding bootstrap support value of the equivalent maximum likelihood distances analysis (see methods). The tree suggests that the *Entamoeba* tryptophanase was acquired by HGT from a relative of the anaerobic bacterial genus *Fusobacterium*. By contrast, the *Trichomonas vaginalis* gene appears to have a separate origin by HGT from a relative of the anaerobic *Bacteroides* group. The *scale bar* represents 10% of inferred sequence divergence. Both the GenBank and RefSeq accession numbers are given for the *Entamoeba* entry. The EC number is also shown.

evolution of the *Trichomonas* and *Entamoeba* genomes and their metabalomes. Mapping the HGT onto a schematic of *Entamoeba* metabolism *(10)* indicates that HGT has affected some important pathways including iron-sulfur cluster biosynthesis, amino acid metabolism, and nucleotide metabolism (**Fig. 28.3**). In *Trichomonas*, metabolic pathways affected include salvage pathways, amino acid metabolism, synthesis of lipophosphoglycan, and many more (**Fig. 28.4**). For *E. histolytica*, 45% of the inferred transferred genes are hypothetical or unclassified proteins and in

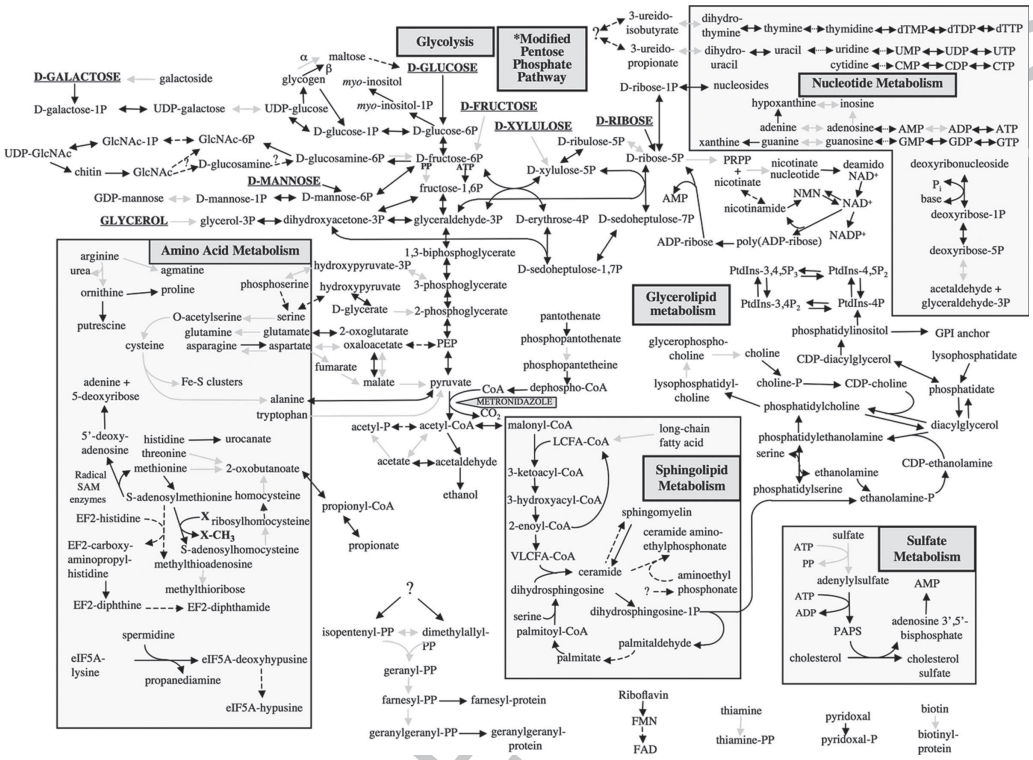

Fig. 28.3. **Predicted metabolic pathways of *E. histolytica* based on the analysis of its genome showing inferred HGT** *(9)*. Glycolysis and fermentation are the major energy generation pathways. *Bold grey arrows* represent enzymes encoded by genes that are among the 96 candidates for HGT into the *E. histolytica* genome. *Broken arrows* indicate enzymes for which no gene could be identified from the genome data, although the activity is thought to be present. The *framed arrow points* to the target of Metronidazole, the major drug for treatment of amoebic liver abscess. Abbreviations: PEP, phosphoenolpyruvate; GlcNAc, *N*-acetylglucosamine; LCFA, long chain fatty acid; VLCFA, very long chain fatty acid; PRPP, phosphoribosyl pyrophosphate; GPI, glycosylphosphatidylinositol; PAPS, phosphoadenosine phosphosulfate.

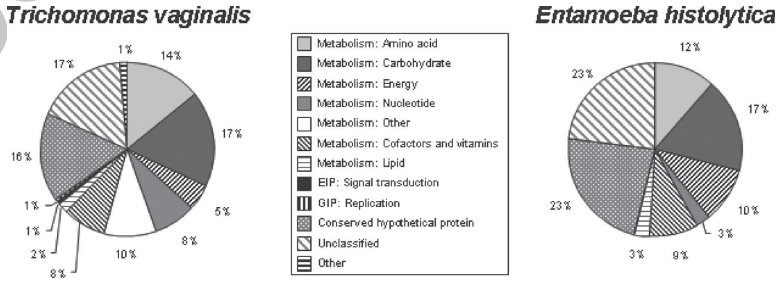

Fig. 28.4. **Functional categories among 153 *Trichomonas vaginalis* and 68 *Entamoeba histolytica* candidate HGT**. Distribution of functional annotation from the KEGG database among candidate HGT. The % values were rounded up.

T. vaginalis the corresponding value is 33%. These values may simply reflect the observation that for *Entamoeba histolytica* and *Trichomonas vaginalis* around 30% of the proteins predicted from the genome sequence are also hypothetical or unclassified.

4. Will Improved Sampling Affect Our Inference of HGT?

4.1. Re-analysis of 96 Putative HGT to Entamoeba histolytica

In order to shed some light on how taxon sampling might affect the inference of HGT, we reanalysed the 96 putative HGTs we identified in the *Entamoeba* genome paper *(10)*, adding data from eukaryotic and prokaryotic genomes published before August 2005 *(27)*. In doing so, we were hoping to gain some insight into how our previous inferences were influenced by the sparse sampling of eukaryotic and prokaryotic genes and species available at the time. Such sparse gene and species sampling is, and is likely to remain, a very serious problem for reconstructing global trees and inferring HGT *(8, 29, 30)*. Thus, although ecologists differ in their claims for the extent of the unsampled microbial world, they all agree that those strains in culture and the even smaller subset for which we have genome data represent the smallest tip of a very large iceberg. Additionally, since this re-analysis coincided with our HGT analysis of the *T. vaginalis* genome *(11)*, taxa and sequences were sampled from the same database, thus facilitating comparison between the two genomes.

A total of 41 HGTs remain as strongly supported as before based upon the original criteria. For the remaining 56 tree topologies, support for recent HGT into the *Entamoeba* lineage is not as strong as before. For 27 of these 56 trees, where previously there were two strongly supported nodes separating *Entamoeba* from other eukaryotes, the hypothesis of HGT is now supported by only one well-supported node. However, close scrutiny of the bootstrap partition tables for these trees revealed that, as before, there are no trees in which *Entamoeba* is found together with another eukaryote. Thus, HGT still remains the best hypothesis to explain 68 (70%) of the original 96 topologies. In a further 14 trees, the position of *Entamoeba* among prokaryotes and eukaryotes was not well supported. The taxonomic sampling of eukaryotes in these trees is now very patchy and the trees do not resemble consensus eukaryotic relationships. Thus, although the trees do not fulfil our original conservative criteria for HGT, they do not provide strong support for the alternative hypothesis that the *Entamoeba* genes were vertically inherited from a common ancestor shared with other eukaryotes.

4.2. Eukaryote-to-Eukaryote Transfers?

In nine of the re-sampled trees, *Entamoeba* either clustered with a single newly published eukaryotic sequence, or we could not reject such a relationship among mainly prokaryotic sequences. Six of these nine trees recovered *Entamoeba* and *Trichomonas* together, and two trees grouped *Entamoeba* with the diatom *Thalassiosira* (for example, *see* **Fig. 28.5**). Such trees are not easy to explain by simple vertical inheritance for the species concerned

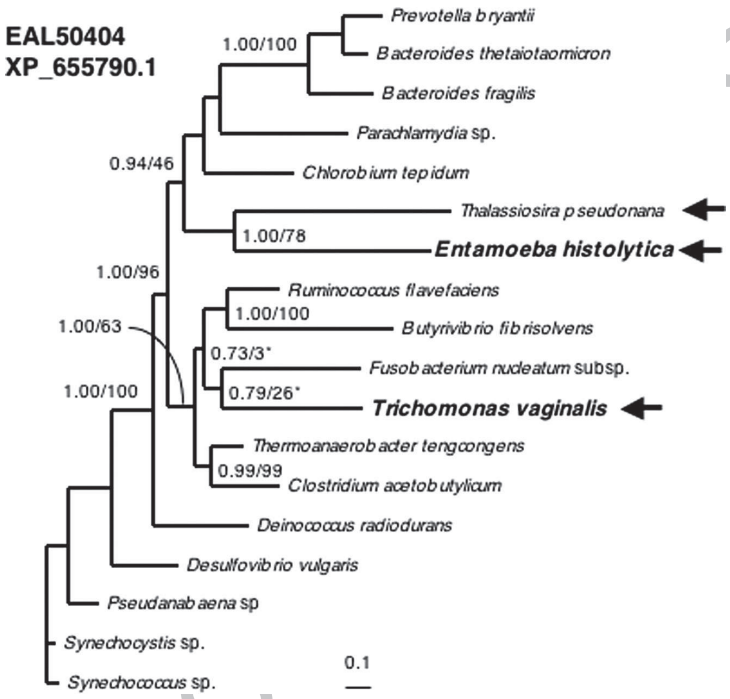

Fig. 28.5. **Some trees show unusual relationships for eukaryotes that are not easy to explain within the framework of the current consensus for eukaryotic phylogeny**. *Entamoeba* and the diatom *Thalassiosira* are closely related in the tree for glutamine synthase (EC 6.3.1.2), but no external data suggest that these two eukaryotes are closely related. One possible explanation is eukaryote-to-eukaryote HGT. *Trichomonas vaginalis* also contains a homologue of glutamine synthase, but in this case it clusters weakly with *Fusobacterium*. The *scale bar* represents 10% of inferred sequence divergence. Both the GenBank and RefSeq accession numbers are given for the *Entamoeba* entry.

within the framework of the current consensus for eukaryotic relationships *(31–33)*. Similar unusual topologies have been previously reported for other eukaryotes *(5)* (compare **Chapter 27**). The explanations advanced to explain the absence of the gene in other eukaryotes, include massive gene loss from eukaryotic lineages, or HGT between the eukaryotes concerned. *Entamoeba* can phagocytose both eukaryotes and prokaryotes, and it has been suggested that HGT between eukaryotes, after one of them had acquired the gene from a prokaryote, could explain such peculiar tree topologies and sparse eukaryotic sampling *(5)*. The fact that six among nine entries recover a relationship between *Entamoeba* and *Trichomonas*, the other mucosal pathogen of humans discussed here, is consistent with this idea. Recent large-scale analyses already support the hypothesis that prokaryotes from the same environment may share a set of niche-specific genes *(34, 35)*.

For five trees, the gene now appears to be present in several eukaryotes from different taxonomic groups and the trees

cannot exclude a common origin for all eukaryotic sequences. Thus, for about 5% of the original trees, the simplest hypothesis to explain the observed pattern for the *Entamoeba* gene is no longer HGT, but vertical inheritance from a common ancestor shared with other eukaryotes.

5. Conclusions

The results of our study demonstrate that HGT has played a significant role in the evolution of the *Trichomonas* and *Entamoeba* genomes. The majority of functionally categorised HGTs are enzymes involved in metabolism, thus affecting various metabolic pathways. This is consistent with previously published continual transfer hypotheses, such as the complexity hypothesis *(27)*, which posit that HGT of genes involved in processing a single substrate are more likely to be transferred than those genes encoding proteins that interact with many other cellular components, such as proteins found in the ribosome. Furthermore, phylogenetic analyses imply that a broad range of donors have contributed the acquired genes with a bias towards prokaryotes that share the same ecological niche as *Entamoeba* and *Trichomonas*. Given that our screen was designed to detect only "recent" HGTs, the cases we record may be only the tip of a much larger historical iceberg *(1)* of transfers. The results are thus consistent with ideas that prokaryote to eukaryote transfers have occurred continually throughout eukaryotic history *(6)*.

Acknowledgments

This work was supported by a Wellcome grant awarded to TME, RPH and UCMA. We would like to thank Dr Simon Harris for advice on HSSP analysis.

References

1. Doolittle, W. F., Boucher, Y., Nesbo, C. L., Douady, C. J., Andersson, J. O., Roger, A. J. (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos Trans R Soc Lond B Biol Sci* **358**, 39–57.
2. Koonin, E. V., Makarova, K. S., Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**, 709–42.
3. Lawrence, J. G. (2005) Horizontal and vertical gene transfer: the life history of pathogens. *Contrib Microbiol* **12**, 255–71.
4. Ochman, H., Lawrence, J. G., Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
5. Andersson, J. O. (2005) Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* **62**, 1182–1197.

6. Doolittle, W. F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trend Genet* **14**, 307–11.
7. Lawrence, J. G. (2005) Common themes in the genome strategies of pathogens. *Curr Opin Genet Dev* **6**, 584–8.
8. Richards, T. A., Hirt, R. P., Williams, B. A., Embley, T. M. (2003) Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**, 17–32.
9. Andersson, J. O., Sjögren, A. M., Horner, D. S., Murphy, C. A., Dyal, P. L., Svärd, S. G., Logsdon, J. M. Jr, Ragan, M. A., Hirt, R. P., Roger, A. J. (2007) A genomic survey of the fish parasite *Spironucleus salmonicida* indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. *BMC Genomics.* **14**, 8–51.
10. Loftus, B., Anderson, I., Davies, R., Alsmark, U. C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M. A., Rabbinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillen, N., Gilchrist, C., Stroup, S. E., Bhattacharya, S., Lohia, A., Foster, P. G., Sicheritz-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N. M., Petri, W. A., Jr., Clark, C. G., Embley, T. M., Barrell, B., Fraser, C. M., Hall, N. (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–8.
11. Carlton, J. M., Hirt, R. P, Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedr, D., Barton, G. J., Westrop, G. D., Müller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbos, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M. Jr., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M., Johnson, P. J. (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*. **315**, 207–12.
12. Nanda, N., Michel, R. G., Kurdgelashvili, G., Wendel, K. A. (2006) Trichomoniasis and its treatment. *Expert Rev Anti Infect Ther* **1**, 125–35.
13. Petri, W. A. Jr., (2000) Protozoan parasites that infect the gastrointestinal tract. *Curr Opin Gastroenterol.* **1**, 18–23.
14. Shoemaker, N. B., Vlakamis, H., Slyers, A. A. (2001) Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl Environ Microbiol* **67**, 561–8.
15. Sicheritz-Ponten, T., Andersson, S. G. (2001) A phylogenomic approach to microbial evolution. *Nucleic Acids Res* **29**, 545–52
16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–402.
17. Thompson, J. D., Higgins, D. G., Gibson T. J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–80.
18. Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540–52
19. Huelsenbeck, J. P., Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–55.
20. Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A., Winka, K. (2003) Comparing bootstrap and posterior probability values in the four-taxon case. *Syst Biol* **52**, 477–87.
21. Foster, P. G. P4. http://www.bmnh.org/~pf/p4.html
22. Desper, R., Gascuel, O. (2004) Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol Biol Evol* **21**, 587–98.
23. Timmis, J. N., Ayeliffe, M. A., Huang, C. Y., Martin,W. (2004) Endosymbiotic gene transfers: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **5**, 123–35.
24. Rost, B., (1999) Twilight zone of protein sequence alignments. *Protein Eng* **12**, 85–94.
25. Nair, R., Rost, B. (2002) Sequence conserved for subcellular localization. *Protein Sci* **11**, 2836–47.
26. Reid, G., Bocking, A. (2003) The potential for probiotics to prevent bacterial vagi-

nosis and preterm labor. *Am J Obstet Gynecol.* **189**, 1202–8.

27. Clark, C. G., Alsmark, U. C. M., Hofer, M., Saito-Nakano,Y., Ali, V., Marion, S., Weber, C., Mukherjee, C., Bruchhaus, I., Tannich, E., Leippe, M., Sicheritz-Ponten, T., Foster, P. G. Samuelson, J., Noël, C. J., Hirt, R. P., Embley, T. M., Gilchrist, C. A., Mann, B. J., Singh, U., Ackers, J. P., Bhattacharya, S., Bhattacharya, A., Lohia, A.,Guillén, N., Duchêne, M., Nozaki, T., Hall, N. (2007) Structure and Content of the *Entamoeba histolytica* Genome. *Adv Parasitol*, **65**, 51–190.
28. Jain, R., Rivera, M. C., Lake, J. A. (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801–6.
29. Andersson, J. O., Doolittle, W. F., Nesbo, C. L. (2001) Genomics. Are there bugs in our genome? *Science* **292**, 1848–50.
30. Salzberg, S. L., White, O., Peterson, J., Eisen, J. A. (2001) Microbial genes in the human genome: lateral transfer or gene loss? *Science* **292**, 1903–6.
31. Baldauf, S. L. (2003) The deep roots of eukaryotes. *Science* **300**, 1703–6.
32. Embley, T. M., Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature.* **440**, 623–30.
33. Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., McCourt, R. M., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Nerad, T. A., Shearer, C. A., Smirnov, A. V., Spiegel, F. W., Taylor, M. F. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* **52**, 399–451.
34. Beiko, R. G., Harlow, T. J., Ragan, M. A. (2005) Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* **102**, 14332–7.
35. Mira, A., Pushker, R., Legault, B. A., Moreira, D., Rodriguez-Valera, F. (2004) Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. *BMC Evol Biol* **4**, 50.

AQ1

Chapter-28

Query No.	Page No.	Line No.	Query
AQ1	500	11	The reference *(28)* is not cited in text part. Please provide.